Unlocking prophage potential: *In silico* and experimental analysis of a novel *Mycobacterium fortuitum* LysinB containing a peptidoglycan-binding domain

Ritam Das[1,2], Kanika Nadar[3], Ritu Arora[3] and Urmi Bajpai[3]*

[1]Faculty of Biological Sciences, Friedrich Schiller University, Jena-07737, Germany

[2]Department of Life Science, Acharya Narendra Dev College, University of Delhi, Govindpuri, Kalkaji-110019, New Delhi, India

[3*]Department of Biomedical Science, Acharya Narendra Dev College, University of Delhi, Govindpuri, Kalkaji-110019, New Delhi, India, urmibajpai@andc.du.ac.in

Address for correspondence:

Urmi Bajpai, PhD (Professor)

Department of Biomedical Science,

Acharya Narendra Dev College (University of Delhi),

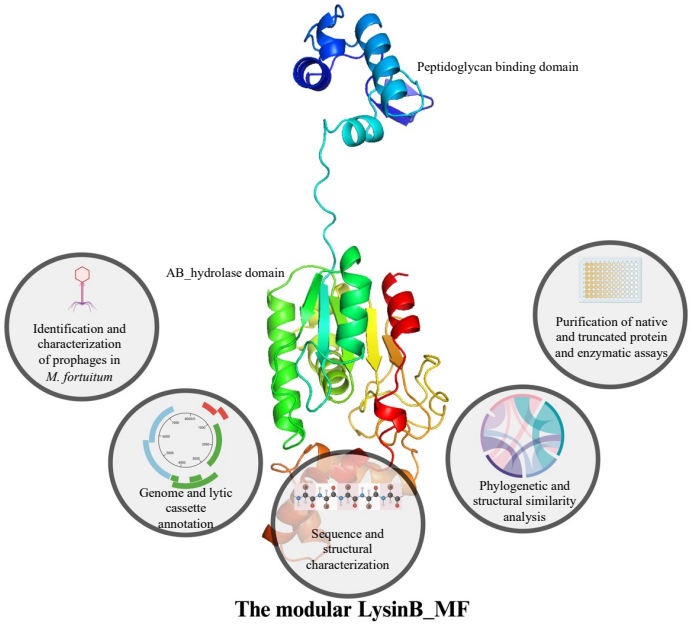
Govindpuri, Kalkaji, New-Delhi-110019, India.

Email id: urmibajpai@andc.du.ac.in

Abstract

Endolysins are highly evolved bacteriophage-encoded lytic enzymes produced to damage the bacterial cell wall for phage progeny release. They offer promising potential as highly specific lytic proteins with a low chance of bacterial resistance. The diversity in lysin sequences and domain organization can be staggering. *In silico* analysis of bacteriophage and prophage genomes can help identify endolysins exhibiting unique features and high antibacterial activity, hence feeding the pipeline of narrow-spectrum protein antibiotics. Mycobacteriophage lysis cassettes mostly have two lytic enzymes, LysinA and LysinB. The enzyme LysinA targets peptidoglycan in the cell wall and possesses a modular architecture. LysinB typically contains a single domain and acts upon the mycolyl ester linkages in mycolyl-arabinogalactan-peptidoglycan (Payne *et al.*, 2010). This study aimed to find novel LysinBs against *Mycobacterium fortuitum*. After a detailed *in silico* characterization of lysis cassettes from three *M. fortuitum* prophages, we chose to work on a LysinB (hereafter described as LysinB_MF) found in an incomplete prophage (phiE1336, 9.4 kb in strain E1336).

LysinB_MF showed low sequence similarity with any other endolysins in the database and formed a separate clade on phylogenetic analysis. LysinB_MF's structure, extracted from the AlphaFold Protein Structure Database, demonstrated a modular architecture with two structurally distinct domains: a peptidoglycan-binding domain (PGBD) at the N-terminal and the characteristic alpha/beta hydrolase domain connected via a linker peptide. We found the alpha/beta hydrolase domain, which is the enzyme-active domain (EAD), contains the conserved Ser-Asp-His catalytic triad with a tunnel-like topology and forms intermolecular hydrogen bonds. The PGBD shows structural similarity to the cell-wall binding domain of an amidase from *Clostridium acetobutylicum,* hinting at its acquisition due to domain mobility. Our *in silico* electrostatic potential analysis suggested that PGBD might be essential to the enzyme activity. This was experimentally validated by generating a truncated version of the enzyme, which demonstrated about six-fold decreased activity compared to its native form. The antimycobacterial activity of this enzyme was also compromised in its absence. Based on our analysis, PGBD emerged as an integral constituent of enzymes with diverse functional properties and is predicted to be a conserved cross-kingdom. Overall, this study highlights the importance of mining mycobacterial prophages as a novel endolysin source. It also provides unique insights into the diverse architecture of mycobacteriophage-encoded endolysins and the importance of functional domains for their catalytic activities.
Keywords: Prophage; Endolysins; LysinB; Structure; Domains; Esterase activity



The modular LysinB_MF

1. Introduction

Mycobacterium infections are among the most debilitating, if not fatal (Gordon & Parish, 2018; Das *et al.*, 2021). *Mycobacterium tuberculosis*, the causative agent of Tuberculosis, alone was the cause of 1.3 million deaths in the year 2022, the second-highest infectious killer after COVID-19 (Bagcchi, 2023). Infections due to nontuberculous *Mycobacterium* (NTM) are also becoming an emerging issue of concern, and their prevalence is increasing worldwide. While *Mycobacterium fortuitum*-associated pulmonary diseases are the most prevalent (Khosravi *et al.*, 2018), the bacterium also causes extra-pulmonary infections in humans, such as cutaneous and bone lesions, which are difficult to treat, chronic, occult, resistant, and recurrent (Sethi *et al.*, 2014; Fraga *et al.*, 2022).

Mycobacteriophages are viruses that specifically infect *Mycobacterium spp*. Most mycobacteriophages code for two endolysins: LysinA and LysinB, which target the bacterial cell wall and cause lysis of their host cells at the end of their life cycle (Payne *et al.*, 2010; Payne & Hatfull, 2012). LysinA is a peptidoglycan hydrolase, and LysinB is a mycolylarabinogalactan esterase, which hydrolyzes the mycolyl ester linkages in mycolyl-arabinogalactan-peptidoglycan

78 layer present in mycobacteria. Since the exogenous application of endolysins has been
79 demonstrated to show an antibacterial effect, lysins have been explored as an alternative/
80 combination therapy with antibiotics in several studies (Fraga *et al.*, 2019; Abouhmad *et al.*, 2020;
81 Hurst-Hess *et al.*, 2023; Singh *et al.*, 2023). While the reports on structural and functional
82 characterization of mycobacteriophage-encoded lysins (Payne *et al.*, 2010; Eniyan *et al.*, 2020;
83 Gigante *et al.*, 2021) have provided insights into their structural diversity, these lysins are relatively
84 less reported when compared to phage lysins against Gram-negative and Gram-positive bacteria.
85 Besides the isolated mycobacteriophages, the genomes of nontuberculous *Mycobacterium* (NTM)
86 that are enriched with prophages (Abad *et al.*, 2023), could also serve as a promising source for
87 diverse antibacterial proteins (Schmitz *et al.*, 2011; Shah *et al.*, 2023). However, these prophages
88 are largely unannotated. In this study, we report a novel LysinB encoded by one such unique
89 prophage in a strain of *M. fortuitum,* with three characteristic features: i) unlike the usual single-
90 domain architecture of LysinB enzymes, LysinB_MF has a structurally distinct peptidoglycan
91 binding domain (PGBD), ii) its enzyme active domain shows structural similarity to a hydrolase
92 (amidase) of *Clostridium acetobutylicum,* and iii) PGBD has an important role in enzyme activity.
93 As a first step towards understanding the molecular properties, we have characterized the lytic
94 cassette of three *M. fortuitum* prophages, and carried out a detailed examination of LysinB_MF,
95 showing the highest similarity with D29 LysinB, which we used as a bait sequence for searching
96 LysinB sequences in *M.fortuitum* prophages. We have carried out an extensive sequence and
97 structural analysis of lysis cassettes and LysinB_MF and hypothesized and biochemically
98 characterized the importance of the PGBD in its activity.
99
100

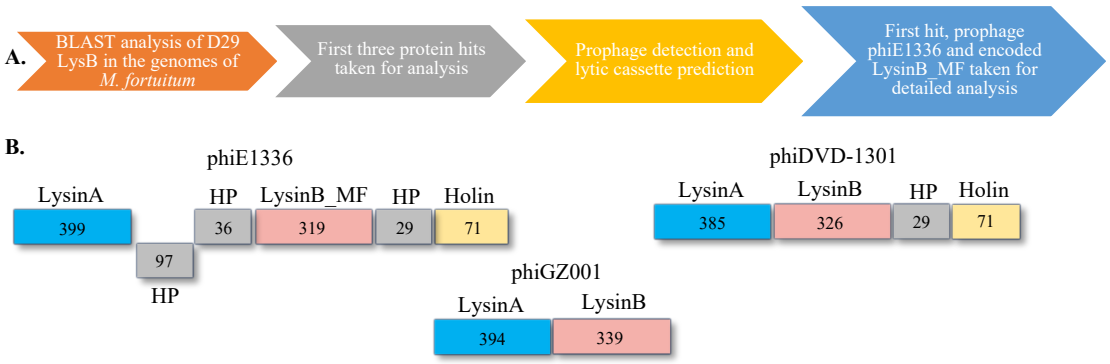
101 **2. Results and Discussion**

102 **2.1 *Mycobacterium* prophages as a source of novel lysin sequences**

103 D29 and Ms6 LysB enzymes are the extensively studied mycolylarabinogalactan esterases with
104 pre-clinical characterization (Payne & Hatfull, 2012; Gigante *et al.*, 2017; Fraga *et al.*, 2019;
105 Hurst-Hess *et al.*, 2023). However, to facilitate the growing interest in mycobacteriophage lysins
106 to treat drug-resistant *Mycobacterium spp.,* more lysins need to be discovered and investigated to
107 understand their biological features and to develop them as novel therapeutics. Non-tubercular
108 mycobacteria (NTM) are reported to harbour a large proportion of prophages (Abad *et al.*, 2023).
109 We used D29 LysB (O64205) as a bait sequence and performed BLAST to find similar but not
110 identical LysinB sequences against all the *M. fortuitum* genomes in the database. The workflow
111 for the *in silico* identification of LysinB enzymes is represented in Figure 1A. The top three hits
112 for LysinB were found to be encoded by prophages of *M. fortuitum* strains (Table 1). In this study,
113 we have carried out the structural and functional characterization of a unique LysinB enzyme
114 (mentioned as LysinB_MF) encoded by a prophage, called phiE1336, present in the sequenced
115 strain E1336 of *Mycobacterium fortuitum* (tax id: 1766). A detailed characterization of this
116 prophage has been done in the later sections.

117 We found the prophage phiE1336 genome to be incomplete (9.4 kb), however, its lysis cassette
118 was intact. It constituted LysinA, three small hypothetical proteins, LysinB and Holin (Figure 1B).
119 LysinB_MF, with an overall identity of 49% with D29 LysinB, is a peptidoglycan-binding domain-
120 containing protein (WP_064866887.1). The second hit, with an overall identity of 48%, was a
121 hypothetical protein (HP) (WP_278289780.1) from *M. fortuitum* strain DVD-1301
122 (NZ_JARUNL010000131.1). Reannotation of the genome and PHASTER prediction indicated the
123 protein to be a LysinB, encoded by an intact prophage (hereby called phiDVD-1301). This

124 prophage genome size was 39.6 kb with a GC content of 62.86%, similar to what is observed for
125 other mycobacteriophages. Nucleotide BLAST analysis of the genome indicated phiDVD-1301 to
126 have approximately 80% sequence identity with cluster P phages Tortellini (KX648391),
127 Phayonce (KR080195) and Purky (MN096355). Annotation demonstrated the presence of a lytic
128 cassette (Figure 1B) with LysinA (385 amino acids) as the largest protein-coding gene. LysinB
129 (326 amino acids) from phiDVD-1301 showed 48% identity with D29 LysB, however, domain
130 predictions for this enzyme remained unsuccessful. The third hit (WP_269975983.1) showing 36%
131 sequence identity with D29 LysB was a hypothetical protein (HP), encoded by an *M. fortuitum*
132 strain MF GZ001 (NZ_CP107719.1) prophage. This prophage (hereby called phiGZ001) has a
133 size of 38.9 kb with distinct attL (CACACTGTTCTTC) and attR (CACACTGTTCTTC) sites
134 recognized by serine integrases (Suzuki *et al.*, 2020). Nucleotide BLAST analysis of the genome
135 indicated phiGZ001 to have approximately 86% sequence identity with cluster F1 phages XFactor
136 (KT281795), Tootsieroll (MZ747519) and Blexus (MN586012). Sequence reannotation and
137 domain analysis of the identified HP (WP_269975983.1) predicted an alpha/beta hydrolase
138 domain. In phiGZ001, we did not find a holin gene in the lysis cassette (Figure 1B). We speculate
139 this prophage either lyses the host independent of holin (Eniyan *et al.*, 2020) or the lytic cassette
140 is incomplete.



141

Figure 1: Analysis of prophages in *M. fortuitum* strains. A) The workflow of prophage detection and characterization of the lytic regions. **B)** Lytic cassette of prophages phiE1336, phiDVD-1301 and phiGZ001.

Table 1: Information on *M. fortuitum* prophages identified and characterized in the study

S. no.	Name	Size (kb)	Completeness	GC%	Cluster similarity	*M. fortuitum* lysogen
1	phiE1336	9.4	Incomplete	62	-	NZ_LZSJ01000108.1
2	phiDVD-1301	39.6	Intact	62.8	P	NZ_JARUNL010000131.1
3	phiGZ001	38.9	Intact	63.4	F1	NZ_CP107719.1

2.2 The incomplete prophage in the *M. fortuitum* strain E1336

PHAST analysis predicted the presence of an incomplete genome of prophage phiE1336. Its size is only 9.4 kb (Figure 2A) but the GC content (62%) is comparable to the host (Ho *et al.*, 2012). Mycobacteriophages are grouped into clusters based on the similarities and differences in their nucleotide sequences (Hatfull, 2022). Currently, 32 clusters of mycobacteriophages have been defined, and the unclassified phages are denoted as singletons (https://phagesdb.org/). We performed nucleotide BLAST analysis to identify the possible cluster of phiE1336 and to predict phages identical to it. None of the whole genome sequences in the database showed more than 10% query coverage, a trend similar to singletons. However, considering the short genomic fragment of phiE1336, it is difficult to state conclusively whether it belongs among the singleton phages or to a cluster.

2.3 The lytic cassette in prophage phiE1336

We extracted the 9.4 kb fragment and performed annotation using RAST, which predicted it to have 13 ORFs (Figure 2A). Among the predicted ORFs, six could be assigned putative functions, which included lytic and structural proteins (Table S1). The lytic cassette begins with LysinA (44.3

167 kDa), a peptidoglycan hydrolase consisting of an amidase domain spanning from 171–317 and an
168 overlapping peptidoglycan recognition protein (PGRP) region from 183–317 amino acids (Figure
169 2B). Co-occurrence of domains is common in phage lysins (Payne & Hatfull, 2012; Hatfull, 2018).
170 The amidase domain targets the N-acetylmuramoyl-L-alanine linkages, which connect the sugar
171 backbone to the peptide crosslink (Payne & Hatfull, 2012). BLAST analysis showed phiE1336
172 LysinA to share 71.9% identity with LysinA (gp24) from F1 sub-cluster mycobacteriophage
173 Koella (MH316564) and 70.5% identity with LysinA (gp69) from O cluster phage Smooch
174 (MN428052). Domain construction of lysins from these phages is similar and constitutes the active
175 site of the amidase with an overlaying peptidoglycan recognition protein (PGRP) region.
176 The two small HPs between LysinA and LysinB of phiE1336 (Figure 1B) are of size 10.6 kDa and
177 3.9 kDa and do not possess a recognizable domain. Neither of the HPs shows any significant
178 similarity to other sequences in the database. This arrangement of the lytic cassette, with HPs
179 between Lysin A and LysinB, appears unique for phiE1336 among the other prophages
180 characterized in this study (Figure 1B). The other reported mycobacterium prophages, such as
181 phiMAV1 and MabR, constitute lytic regions with LysinA and LysinB present together (Payne *et*
182 *al.*, 2010; Cote *et al.*, 2022).
183 Upstream to the two HPs in the cassette are the genes encoding for mycolylarabinogalactan
184 esterase (LysinB_MF), another small HP (3.1 kDa), and holin (Figure 2B). LysinB_MF is a
185 multidomain enzyme with a molecular weight of 35.3 kDa. While we could not precisely predict
186 the function of the 3.1 kDa protein, BLAST analysis showed similar proteins to have an ATP
187 binding activity. The holin gene encodes for a protein of size 7.4 kDa, which shares 61.1% identity
188 with the holin from cluster I2 mycobacteriophage Che9c. Comparable to other phage-encoded

holin, this, too, consists of two transmembrane helices, which could destabilize the membrane proton motive force of the host in a time-dependent manner (Pollenz *et al.*, 2022).

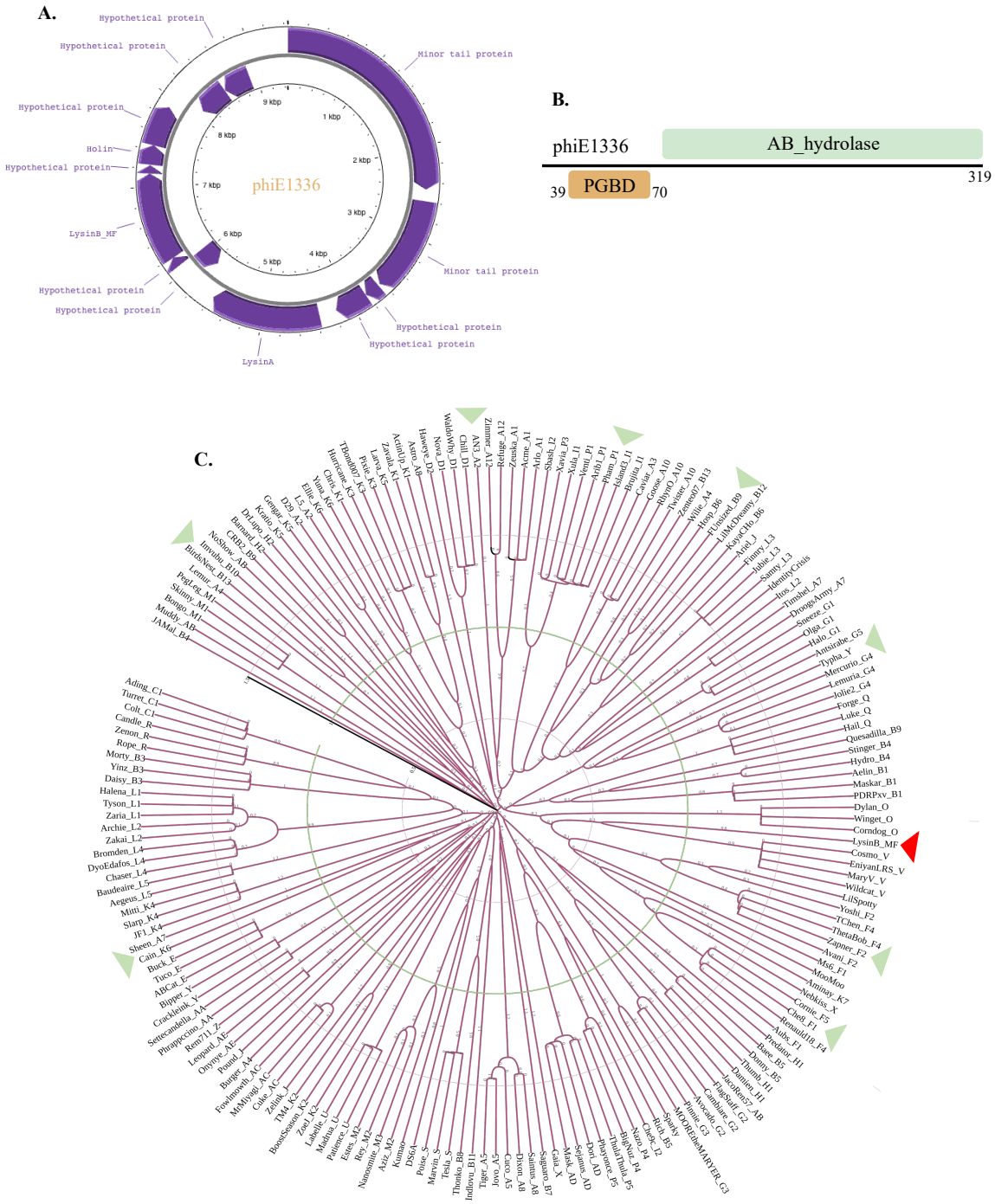


Figure 2: Characterization of phiE1366 and the encoded LysinB_MF. A) phiE1336 has a genome size of 9.4 kb with a GC content of 62%. The genome consists of 13 ORFs, which encode for a few structural proteins and proteins of the lytic cassette **B)** Domain architecture of LysinB_MF from

195 phiE1336 demonstrates distinct PGBD and EAD. **C)** Phylogenetic analysis of LysinB_MF to understand
196 its evolutionary relationship. Representative sequences of LysinBs from all sub-clusters and singletons were
197 taken for analysis. Phylogeny analysis displayed LysinB_MF forming its separate clade (in red arrow), which
198 branched from the neighbouring Cluster V LysinB at some point. Green arrows show the clades disparate with
199 the evolutionary relationships of mycobacteriophage clusters.
200

201 **2.4 LysinB_MF**

202 Although the information on mycobacteriophage-encoded mycolylarabinogalactan esterase
203 (LysinB) has been discussed earlier in several studies (Payne *et al.*, 2010; Korany *et al.*, 2019;
204 Abouhmad *et al.*, 2020), reports on the characterization of a multidomain LysinB is limited
205 (Gigante *et al.*, 2021). We found LysinB_MF to be a modular enzyme composed of two
206 structurally distinct domains connected by a linker peptide (Figure 2B). The enzyme (319 amino
207 acids) is comprised of two functionally defined regions: PGBD and the alpha/beta hydrolase
208 domain (Figure S1). The PGBD is present at the N-terminal region of LysinB_MF and spans from
209 amino acids 39–70, and a linker of 14 amino acids connects it to the catalytic enzyme-active
210 domain (EAD) (Figure 2B). The EAD, like most mycolylarabinogalactan esterases, is comprised
211 of an alpha/beta hydrolase domain that spans right after the linker until the C-terminal end,
212 covering a substantial part of the enzyme. Through BLAST analysis, we found LysinB from a V
213 cluster mycobacteriophage Cosmo (KP027195) to be the closest predicted homologue of
214 LysinB_MF. The sequences showed a high query coverage of 98% and 56.43% identity (e-value
215 3e-116). Besides this, we could not find duplicate or identical proteins to LysinB_MF in the
216 database. Therefore, to examine the evolutionary relationship of LysinB_MF, we performed the
217 phylogenetic analysis using LysinB sequences from singletons and mycobacteriophages belonging
218 to all the sub-clusters (Figure 2C). There were 187 sequences in the repository file prepared with
219 at least one sequence and a maximum of three. The multi-FASTA file was subjected to sequence

220 alignment and then processed using the iTOL to generate a rooted phylogenetic tree based on the
221 maximum likelihood method.
222 Interestingly, we found a separate clade for LysinB_MF, as observed for LysinBs from singleton
223 phages (Figure 2C). The LysinB_MF sequence appeared to have branched from the cluster V
224 endolysins and had a branch length of 0.79. A PGBD in these LysinBs remains unannotated or
225 lacking, although conserved residues in the N-terminal region were observed in the multiple
226 sequence file. It appears Cluster V LysinBs and LysinB_MF may have had a common ancestor at
227 some point; however, they diverged into forming their clade, and LysinB_MF retained the PGBD
228 function.
229 Though the phylogenetic analysis gave us some clue about the evolutionary relationship of
230 LysinB_MF, what was interesting to note was the unconventional grouping of endolysins from
231 different sub-clusters in the same clade (shown in green arrows, Figure 2C). Phylogenetic analysis
232 of tape-measure proteins in mycobacteriophage genomes distinctly shows separate clades, called
233 clusters or even specific sub-clusters (Jacobs-Sera *et al.*, 2012). Phages belonging to the same
234 cluster are grouped into the same clade. However, while studying LysinB, we observed sequences
235 belonging to entirely different clusters, if not sub-clusters, forming a common sub-clade and
236 demonstrating the possibility of a common ancestor at some point in the evolutionary history.
237 LysinB from clusters A7 and K6; F1, F4 and F5; B6 and B12; P1, P3 and I1, respectively, were
238 present in the same clade (Figure 2C). In some instances, the evolution of LysinB appeared to be
239 independent of the evolution of mycobacteriophages. Mycobacteriophage-encoded LysinBs,
240 therefore, emerge as enzymes with a complex evolutionary pattern, and their lineage requires a
241 greater analysis.
242

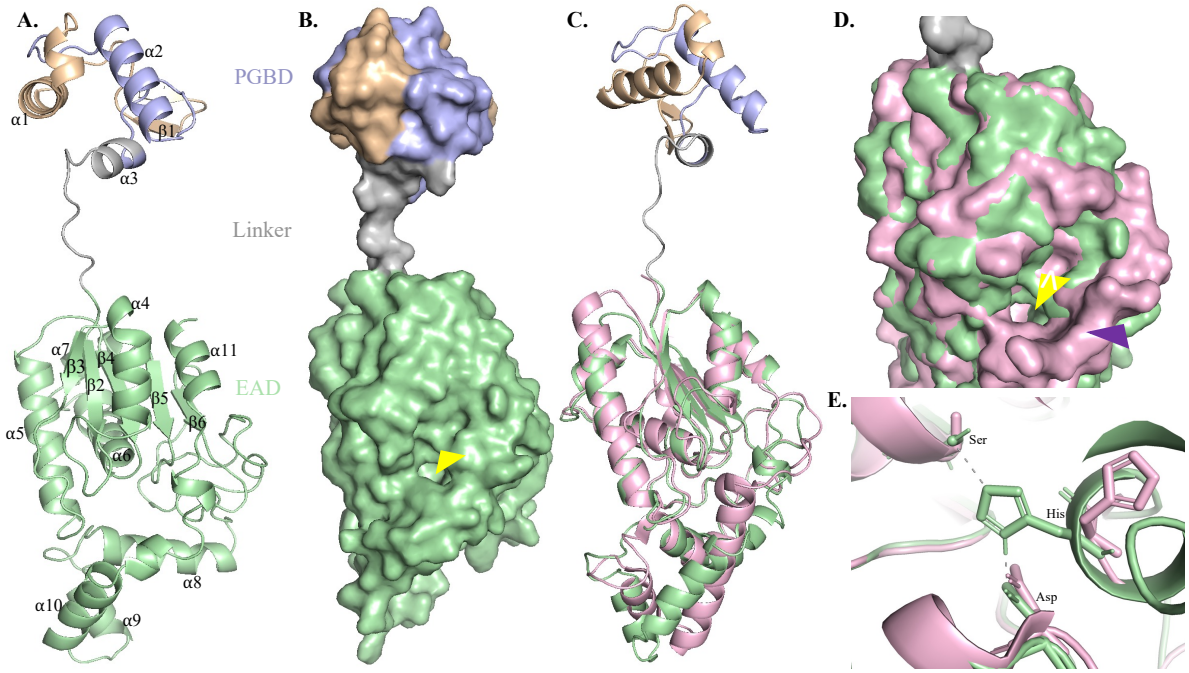


Figure 3: Structural analysis of LysinB_MF. A) Three-dimensional (3D) model of LysinB_MF demonstrates a multi-domain structure comprising two structurally distinct domains: the PGBD (39-70), a short linker region followed by the EAD, an α/β hydrolase (85-317). **B)** The surface model of LysinB_MF demonstrates a bulky C-terminal end with a distinct active-site groove formed with residues Ser-Asp-His. The groove of the EAD, indicated with a yellow arrow, displays a tunnel-like topology. **C)** LysinB_MF can be superimposed on D29 LysB (in pink) with an overall RMSD of 0.56. Structural alignment of the two endolysins shows similar active-site arrangement, although PGBD in D29 Lysin B is absent. **D)** Active-site conformation in LysinB_MF and D29 LysinB. LysinB_MF shows a narrow but deeper topology (shown with a yellow arrow). D29 LysB displays a larger edge, having a funnel-like conformation (shown with a purple arrow). **E)** Active site residues in LysinB_MF (Ser156-Asp236-His305) and D29 LysB (S82-Asp166-His240).

2.5 LysinB_MF has a complex enzyme active domain (EAD)

To analyze the structural information of LysinB_MF, we attempted to characterize the 3D model of the enzyme available in the AlphaFold Protein Structure Database (AFDB: A0A1A0NU50). The structure of this enzyme was available as a peptidoglycan-binding protein derived from the *M. fortuitum* strain described earlier. The protein structure showed a high model confidence (pLDDT > 90) and was taken forward for further analysis (Figure 3).

262 The EAD of LysinB_MF comprises the α/β sandwich belonging to the α/β hydrolase family and
263 contains the catalytic triad Ser-Asp-His. This domain spanned from amino acid 85–317, covering
264 a major portion of the enzyme and demonstrated a highly twisted structure, consisting of a C-
265 terminal linker region from 246-310. The active site of LysinB_MF displayed four β-strands (β2–
266 β5), surrounded from all sides by α-helices (α4-α6). A surface model of the enzyme demonstrated
267 the active-site groove with a tunnel-like topology (Figure 3B). A similar topology is observed in
268 D29 LysB, which serves as a binding pocket for substrates such as several *p*-NP derivatives,
269 trehalose monomycolate and mAGP from mycobacterial cell wall (Payne *et al.*, 2010). A DALI
270 search with high confidence (DALI Z-score 29.7) identified the structural relationship of
271 LysinB_MF with the tertiary structure of D29 LysB (PDB accession 3HC7). The two esterases
272 could be superimposed with an overall RMSD of 0.56 (153 C^{α} atoms) (Figure 3C). Both the
273 structures displayed similar β-sheet conformation around the active site surrounded by the helices.
274 Given the high structural similarity observed between LysinB_MF and D29 LysinB, the topology
275 of D29 LysB's binding pocket, unlike LysinB_MF, showed an edge with a wider circumference
276 and a funnel-like appearance (Figure 3D). Furthermore, a bird's eye view and analysis of the
277 conserved active-site residues (Ser82, Asp166 & His240 in D29 LysB and Ser156, Asp236 &
278 His305 in LysinB_MF) (Figure 4A and 4D) of these enzymes displayed notable differences.
279 Unlike their counterparts, the residues Asp166 in D29 LysB and His305 in LysinB_MF were part
280 of alpha-helices in these enzymes, which added to the differences in their groove topology. Also,
281 unlike D29 LysB, the active-site residue His305 in LysinB_MF was found to form polar contacts
282 with residues Ser156 and Asp236 (Figure 3E). These bonds are known to provide stability by
283 acting as electrostatic attractive forces which influence the charge redistribution in the enzyme and
284 are responsible for the uptake of electrons (pull) from the active site towards the amino acid

residues (Qiu *et al.*, 2018). Also, in the case of serine proteases with similar Ser-Asp-His catalytic triads, the removal of hydrogen bonds resulted in the partial release of the active-site residues in molecular dynamic simulations and a decrease in the overall stability of the transition states, which affects the enzyme's turnover number (k_{cat}) (Zheng *et al.*, 2006).

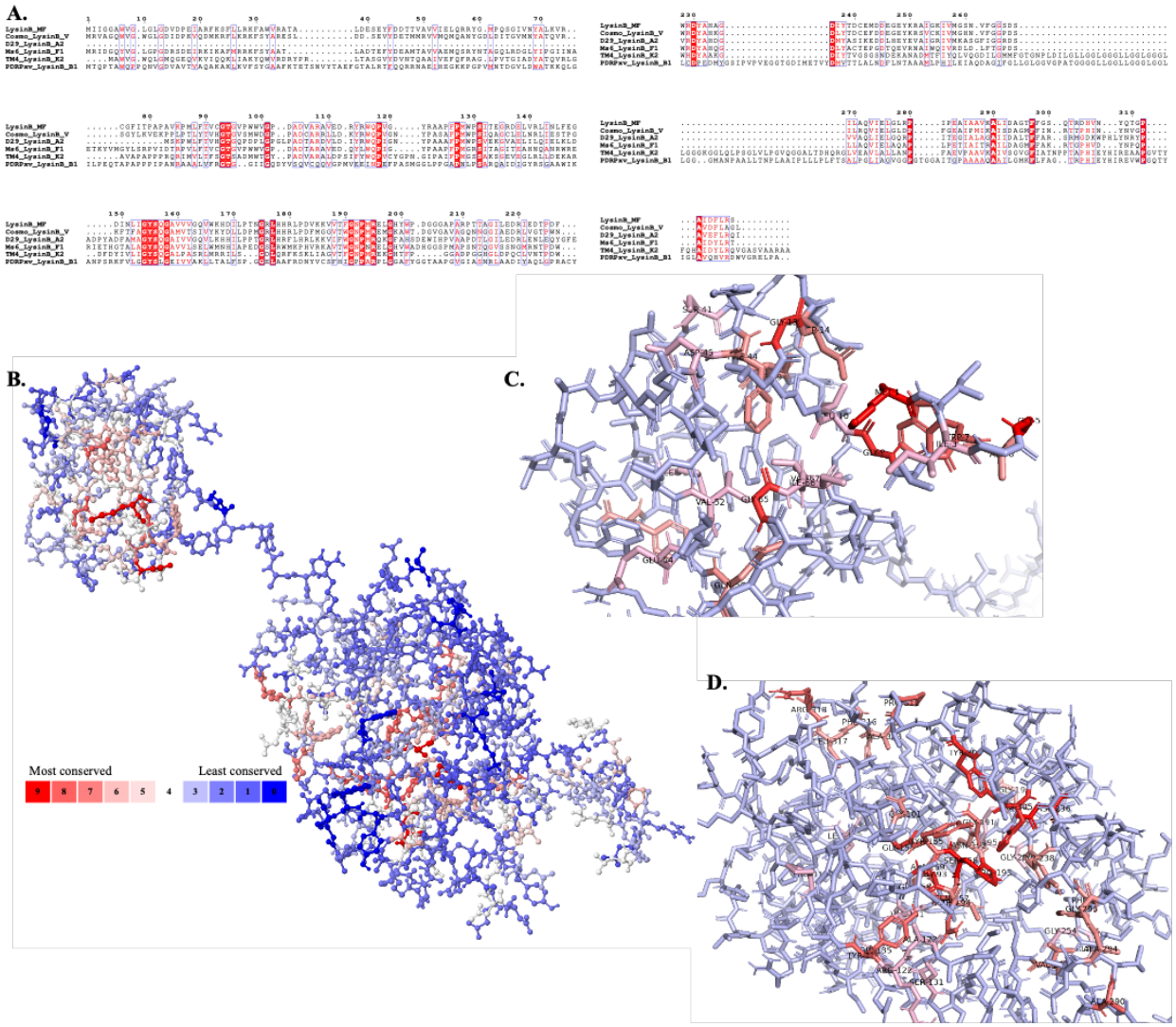


Figure 4: Conserved residues in LysinB_MF. A) Multiple sequence alignment of LysinB enzymes from various clusters demonstrate key residues in PGBD and EAD. The EAD displays the conserved active-site residues Ser-Asp-His and the GXP motif. **B)** Analysis using POLYVIEW-MM identified and mapped conserved residues in the tertiary structure of LysinB_MF. A scale of 0-9 is used to quantify the conservation index, where 0 indicates the least conserved and nine is the most. Exposed residues in the structure of LysinB_MF are found to be least conserved among all others, while most of the highly conserved residues are found buried. **C)** Bird's eye view of PGBD displays conserved amino acids. Ser41, Tyr43 Phe44, Asp45, Val52, Glu54, Gln56, Gly66 and Ile67 are a few highly conserved residues in this domain found both in sequence alignment and in structural

298 analysis. **D)** Bird's eye view of EAD displays conserved amino acids. The active site Ser156, Asp236 & His305
299 and Gly191 & Pro193 from the GXP motif were highly conserved in LysinBs.
300

301

302 **2.6 The PGBD of LysinB_MF is structurally similar to an amidase**

303 In most of the mycolylarabinogalactan esterases reported earlier (Fraga *et al.*, 2019; Abouhmad *et*
304 *al.*, 2020; Eniyan *et al.*, 2020; Singh *et al.*, 2023), an extended N-terminal region consisting of a
305 PGBD hasn't been reported. It is a rare feature that has been observed in Ms6 LysinB (Gigante *et*
306 *al.*, 2021) and our analysis has predicted only seven out of sixty LysinB sequences to contain
307 PGBD (unpublished data). Region 7–71 in LysinB_MF was interpreted as a PGBD-like
308 superfamily, with region 39–70 predicted to be the functional PGBD (highlighted in light blue in
309 Figure 2A). Only a single β-sheet in this region is present in the enzyme's N-terminal end, and the
310 rest of the structure appeared to be highly twisted, dominated by three α-helices. The annotated
311 PGBD is comprised of two α helices (α2 and α3), and it tapers to form the linker region connecting
312 to the EAD. The multiple sequence alignment and structural analysis of LysinB_MF demonstrated
313 several conserved residues in this region (Figure 4), which are also found in LysinBs without the
314 N-terminal PGBD. The amino acids with the highest score in the conservation index are mostly
315 buried in the enzyme and could be involved in the functional properties of the PGBD (Figures 4A-
316 4C).

317 Considering the unavailability of structural information about the PGBD in a mycobacteriophage-
318 encoded LysinB and failure to acquire the same through structural similarity analysis based on the
319 native structure of LysinB_MF, we separately built the PGBD and analyzed it further. The tertiary
320 structure of PGBD was prepared using LysinB_MF as a template, and it had a QMEANDisCo
321 Global score of 0.64 ± 0.11, which ensured high quality (Figure S2A). We then performed
322 structural similarity prediction using DALI, which predicted 20 possible neighbours, and these

323 were visually sorted based on their Z-score. Interestingly, the top scoring structure (Z-score from
324 DALI 8.9) found similar to the PGBD of LysinB_MF was an enzyme with amidase activity (PDB
325 accession code 4XXT). This zinc-dependent amidase is a PGBD-containing protein from
326 *Clostridium acetobutylicum* (ATCC 824). To analyze the structural similarity to PGBD of
327 LysinB_MF, we downloaded the protein 3D file from the PDB database and carried out structural
328 alignment using PyMOL. The PGBDs from the amidase and LysinB_MF could be superpositioned
329 with an overall RMSD as low as 0.46, even though the sequences do not show identical residues
330 (Figure S2B). Also, amidases (or lysozyme-like domains) possessing a peptidoglycan-binding and
331 hydrolyzing function are usually the features of LysinA enzymes. Due to the unavailability of the
332 3D structure of a LysinA in the PDB database, we could not examine the structural similarity of
333 PGBD specifically to a mycobacteriophage-encoded N-acetylmuramoyl-L-alanine amidase.
334 However, it is noteworthy that structural similarity of PGBD of LysinB_MF with high confidence
335 was achieved with an amidase from a bacterial species. We speculate that domain mobility could
336 be a reason behind the generation of modular proteins, and they play a crucial role in their
337 functioning (Basu *et al.*, 2009).
338
339 **2.7 PGBD is essential for the esterase activity of LysinB_MF**
340 To understand the role of PGBD in the activity of LysinB_MF, we first computationally generated
341 its truncated version (LysinB_MF**Δ**PGBD), lacking the N-terminal domain (Figures 5A and 5B).
342 Using eF-surf, a comparative analysis of the native and the truncated structures indicated a change
343 in the electrostatic potential of the active site. The catalytic tunnel appeared more electronegative
344 in LysinB_MF**Δ**PGBD than its native structure (Figures 5C and 5D). A change in electrostatic
345 potential could impact the interactions between an enzyme and its substrate and, hence, the

346 enzyme's activity. The importance of electrostatics in enzyme-mediated catalysis has been
347 demonstrated for amylases where a change in the charge near the active site residues resulted in
348 an enzyme with drastically reduced activity (Nielsen *et al.*, 1999). To find out if the deletion of
349 PGBD also hindered the *in vitro* activity of LysinB_MF, we purified the complete (LysinB_MF)
350 and the truncated enzyme (LysinB_MF**Δ**PGBD) as recombinant proteins (Figures 5E-5F).
351 The lipolytic assay of the truncated protein yielded low Ca^{2+} precipitation (Figure 5H). The *in vitro*
352 esterase assay also revealed the truncated LysinB_MF**Δ**PGBD (Figures 5A, 5B and 5E) to show
353 about a six-fold decrease in activity when compared with the activity of the whole enzyme (Figure
354 5G), indicating PGBD's role in enzyme activity. A similar effect has also been observed for an N-
355 acetylmuramoyl-L-alanine amidase from bacteriophage PSA, where truncation of the cell wall
356 binding domain led to a several-fold decrease in enzyme activity (Korndorfer *et al.*, 2006). The
357 stabilizing effect of PGBD in both mycolylarabinogalactan esterase (LysinB) and N-
358 acetylmuramoyl-L-alanine amidase (LysinA) is significant and requires more investigation.
359

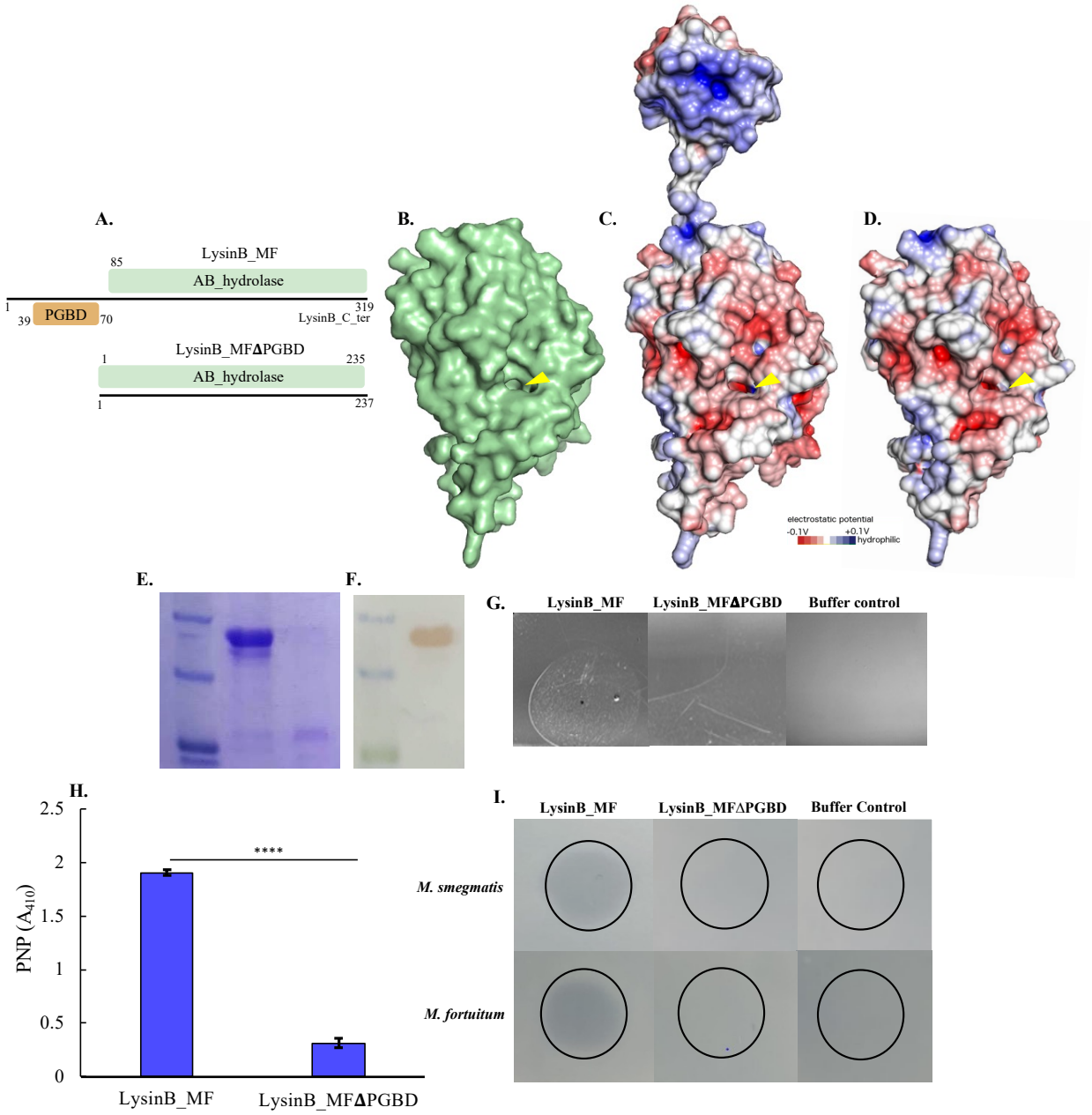


Figure 5: PGBD is essential for the optimum activity of LysinB_MF. A) Domain architecture and comparison of native LysinB_MF and truncated LysinB_MF**Δ**PGBD (EAD). **B)** The structure of LysinB_MF without the PGBD (LysinB_MF**Δ**PGBD) was built using the native structure as a template. LysinB_MF**Δ**PGBD structure comprises the linker region (in grey) and the EAD (in green). **C-D)** Comparison of the intact LysinB_MF and LysinB_MF**Δ**PGBD demonstrates a change in electrostatic potential in the active site (indicated with a yellow arrow). **E)** SDS-PAGE (12%) analysis of purified LysinB_MF and LysinB_MF**Δ**PGBD. Lane 1: protein molecular mass standards. Lane 2: purified recombinant protein LysinB_MF (36.3kDa). Lane 3: LysinB_MF**Δ**PGBD (26 kDa). **F)** Immunoblot of N-terminal His-tagged LysinB_MF. Lane 1: protein molecular mass standards. Lane 2: N-terminal His-tagged LysinB_MF. **G)** A white zone indicates Ca^{2+}-precipitation by LysinB_MF and LysinB_MF**Δ**PGBD (5 µg each). **H)** Esterase

372 activity of whole LysinB_MF and LysinB, determined by the PNP release assay (n=6). **I)** Spot test
373 for intrinsic activity of LysinB_MF and LysinB_MF**Δ**PGBD against *M. smegmatis* and *M.*
374 *fortuitum***.** All the assays were performed using 5 µg of each enzyme preparation.
375

376 **2.8 Antimycobacterial activity of LysinB_MF and LysinB_MF**Δ**PGBD**

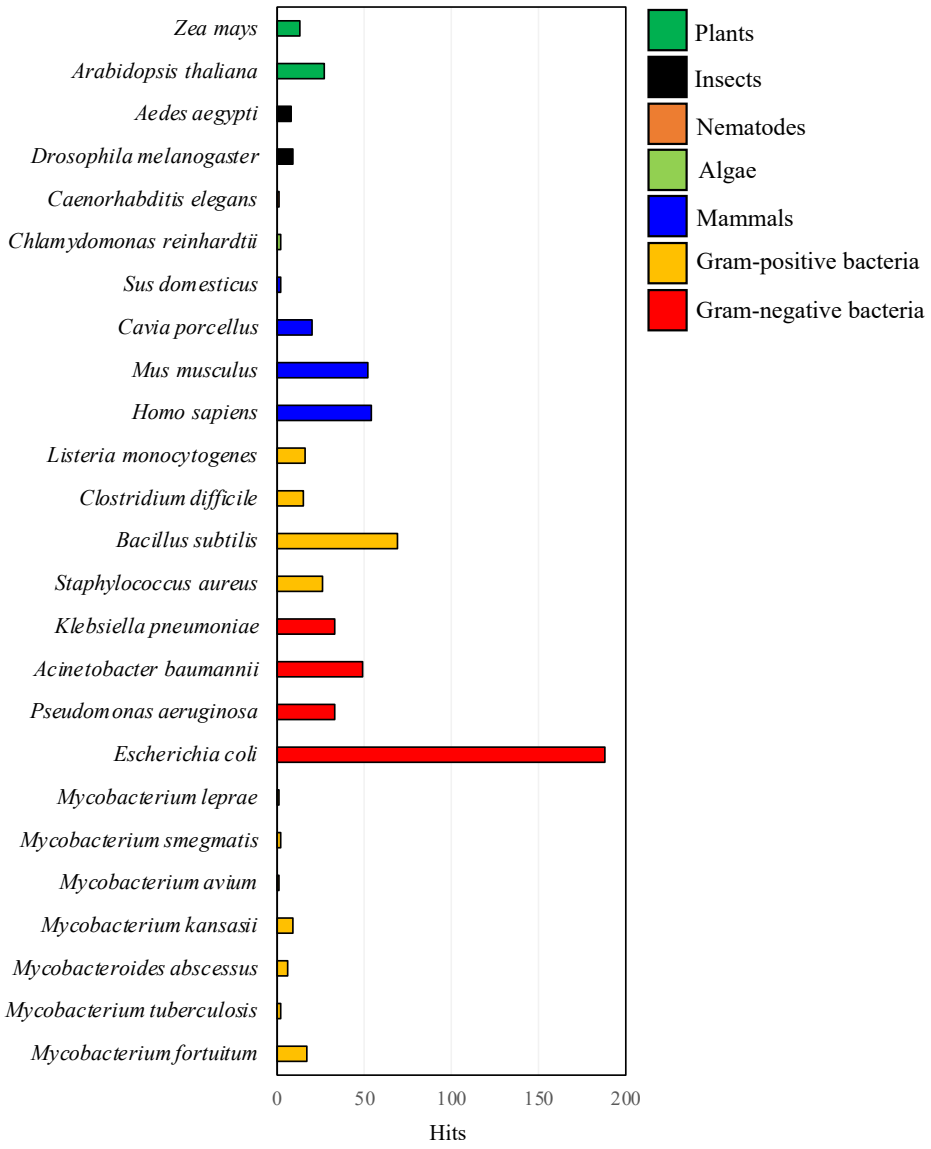
377 Given the complex, hydrophobic cell envelope of mycobacterial spp., the ability of an endolysin
378 to cause cell lysis on external application is highly desired since it obviates the need for an outer
379 membrane permeabilizer (OMP) or a membrane-destabilizer. This property also makes lysins
380 more efficacious candidates as antibiotics (enzybiotics) or as an adjunct to antibiotics. To check
381 whether LysinB_MF possessed intrinsic cell lysis activity, we performed a lysin susceptibility test
382 on *M. fortuitum* and *M. smegmatis*. A zone of clearance on both the mycobacterium species was
383 observed, indicating its ability to lyse mycobacterial cells unaided. Considering the significant
384 decrease in LysinB_MF**Δ**PGBD activity observed in the esterase assay (Figure 5H), we tested its
385 cell lysis activity against *M. fortuitum* (and *M. smegmatis*) and, unlike whole LysinB_MF, did not
386 find clearance zone on the bacterial lawn (Figures 5I). It appears that the PGBD not only supports
387 the enzymatic activity of LysinB_MF but might also have a role in destabilizing the cell wall of
388 *M.fortuitum* and *M.smegmatis*. The cross-species activity of the PGBD has not been reported for
389 a mycobacteriophage lysin. It will be exciting to identify the PGBD-interacting region in the
390 mycobacterial cell wall.
391

392 **2.9 Conservation of PGBD across different kingdoms**

393 Through protein truncation and enzymatic assays, we demonstrated the integral role of PGBD in
394 LysinB_MF activity (Figures 5H- 5I). However, this feature is not constant in all other endolysins.
395 While in the case of bacteriophage PSA, removing the cell-wall binding domain of endolysin
396 demonstrated a similar decrease in enzymatic activity (Korndorfer *et al.*, 2006), the lytic activity

397 of a *Bacillus anthracis* prophage endolysin was enhanced when the C-terminal cell-wall binding
398 domain was removed (Low *et al.*, 2005). PGBD appears to have diverse functionality; however, a
399 proper understanding of the conservation and importance of PGBD is limited. To comprehend the
400 importance of this functional domain in other organisms, we analyzed all the sequences present in
401 the UniProt database against the Pfam entry for the PGBD (PF01471). There are approximately
402 122,000 protein sequences consisting of this domain. From our curated analysis of these sequences
403 based on TaxId from different organisms, we found this domain is conserved across all kingdoms
404 (Figure 6). Among the prokaryotes, PF01471 was predicted to be the highest in *E. coli* (Figure 6),
405 where it is present in LD-transpeptidases, which are involved in cell wall adaptation to stress and
406 for providing outer membrane stability (Aliashkevich & Cava, 2022). In *M. fortuitum*, PF01471
407 was found chiefly in enzymes with a PG hydrolase activity. In the evolutionary Tree of Life in
408 higher kingdoms, with *Arabidopsis thaliana* as an example, our curated search predicted 27
409 UniProt protein matches against PF01471 (Figure 6). Most of the matched proteins in this species
410 were found to be zinc-dependent endo-peptidases or matrix metalloproteinases, which are involved
411 in root development and water uptake (Mishra *et al.*, 2021). In higher animals, including humans,
412 we received 54 UniProt protein matches against PF01471, which, notably, was higher than most
413 prokaryotes analyzed in this study (Figure 6). In humans, a diverse group of essential enzymes
414 possess a PGBD, including matrix metalloproteinases (also observed in *A. thaliana*), enzymes
415 involved in the degradation of extracellular matrix and development of human bones such as
416 Stromelysin-2 (Bord *et al.*, 1998) and collagenases, which are implicated in the turnover of
417 connective tissue matrix constituents (Knauper *et al.*, 1996). The cross-kingdom conservation of
418 the short PGBD observed here would require a deeper analysis and is beyond the scope of our
419 study.

420



421

422 **Figure 6: Cross-kingdom conservation of a PGBD.** The UniProt database was screened against
423 the Pfam ID of PGBD (PF01471), which resulted in approximately 122,000 protein hits with this
424 domain. The curated database of hits was then individually screened against TaxIDs from various
425 organisms across all kingdoms.
426
427
428
429 **3. Conclusion**

430 We have carried out an *in silico* analysis of three prophages in *M. fortuitum*, a medically important

431 NTM. The analysis includes extraction, annotation and sequence characterization of prophages,
432 followed by the structural analysis of the lysis cassettes. In this study, a detailed analysis of one of
433 the three prophages, 'phiE1336' and its encoded LysinB (LysinB_MF) was done. Besides the
434 catalytic alpha/beta domain and a C-terminal linker region, LysinB_MF has an extended N-
435 terminal, containing a peptidoglycan-binding domain. A conservation analysis predicted this
436 domain in diverse enzymes encoded by several organisms. Though LysinB_MF is encoded by an
437 incomplete prophage, which should be incapable of forming phage progeny on induction, its *in*
438 *vitro* esterase activity is greater than the reported LysinB enzymes. It will be worthwhile to
439 investigate whether this highly active, prophage-derived LysinB enzyme has a role in the host
440 *M.fortuitum* life cycle. Based on our *in silico* analysis, we predicted PGBD's role in the enzyme's
441 activity and validated it experimentally. In addition, we also observed a loss of the intrinsic
442 antibacterial activity in the truncated enzyme.
443
444 In summary, the diversity in domain organization of lysins in the prophage-encoded lysis cassettes
445 and the efficient *in vitro* and antimycobacterial activity of LysinB_MF highlights the importance
446 of prophages as a source of novel efficacious endolysins, which could be developed as promising
447 anti-mycobacterial protein antibiotics.
448
449 **4. Materials and Methods**
450 **4.1 *In silico* identification of LysinB_MF and characterization**
451 Lysin B from D29 (GenBank: AAC18452.1) was blasted (Altschul *et al.*, 1990) against all the
452 genomes available in the NCBI database for *M. fortuitum* (tax id: 1766), possibly harbouring
453 prophages. The *M. fortuitum* genome with the top lysin hits (LysinB_MF) was subjected to

454 PHAST (Zhou *et al.*, 2011) and to annotate and fix frameshifts in these genomes, the RAST server
455 was used. The CGview tool (Grant & Stothard, 2008) was used to build the genome map of the
456 first hit, phiE1336. Cluster-wise sequences of LysinB were downloaded from the
457 Actinobacteriophage Database (https://phagesdb.org/), and domains of those lysins were predicted
458 using InterProScan 5 (Quevillon *et al.*, 2005). ClustalW (Thompson *et al.*, 1994) and ESPript
459 (https://espript.ibcp.fr) (Robert & Gouet, 2014) were used to generate alignment files which were
460 then used to create phylogenetic trees using Interactive Tree Of Life (iTOL) v5 (Letunic & Bork,
461 2021). The tertiary structure of LysinB_MF was downloaded from the AlphaFold Protein Structure
462 Database (Varadi *et al.*, 2023) and visualized using PyMOL (https://pymol.org/2/). The
463 POLYVIEW-MM tool was used to predict conserved residues in the tertiary structure of
464 LysinB_MF (Porollo & Meller, 2010). The DALI server was used to find structures similar to the
465 LysinB_MF (Holm, 2022) and eF-surf was used to predict the electrostatic potential of the enzyme
466 (Nielsen *et al.*, 1999). Conservation analysis was done by analyzing the Pfam entry of PGBD
467 (PF01471) against all the proteins available in the UniProt database (https://www.uniprot.org/).
468 TaxIDs were used to detect the presence of this domain across several species using the InterPro
469 classification of protein family website (https://www.ebi.ac.uk/interpro/).
470

471 **4.2 Cloning, overexpression and purification of recombinant LysinB_MF and**
472 **LysinB_MFΔPGBD**

473 The LysinB_MF gene was synthesized and cloned in the pET28a vector by Biomatik Corporation,
474 Kitchener, Ontario. To clone the truncated version of LysinB_MF lacking the PGBD
475 (LysinB_MFΔPGBD), internal primers were designed to PCR amplify the EAD and the amplicon
476 was ligated to the pET28a vector. The recombinant constructs pET28a-LysinB_MF and pET28a-

477 LysinB_MF**Δ**PGBD were overexpressed in BL21(DE3) and C41(DE3) strains of *E. coli,*
478 respectively. For soluble expression, induction with 0.1 mM IPTG and overnight incubation at
479 16ºC was done. The induced culture pellet was collected by centrifugation at 8,000 rpm for ten
480 minutes at 4ºC and sonicated at 90% amplitude with 30 seconds each on or off pulse for ten cycles
481 (Eniyan *et al.*, 2020; Das & Bajpai, 2023). The cell lysate was centrifuged at 13,000 rpm for 45
482 minutes at 4ºC, and the supernatant was considered the soluble fraction.

483 Recombinant LysinB_MF and LysinB_MF**Δ**PGBD were purified by Ni-NTA affinity
484 chromatography. Briefly, the soluble fraction containing the His_6-tagged protein was kept for
485 binding with the Ni-NTA matrix for 2 hours at 4 ºC, and the bound resin was packed in a column.
486 After packing and discarding the flow-through, the column was gradient-washed with buffer
487 containing 20-40 mM imidazole and the bound protein was eluted at 200-250 mM imidazole
488 concentrations. The eluted fractions were analyzed using 12% SDS-PAGE (Laemmli, 1970) and
489 the purified proteins were dialyzed with a buffer containing 50 mM Tris (pH 7.9) and 150 mM
490 NaCl. The concentration of each protein was estimated by Bradford assay (Bradford, 1976). For
491 immunoblotting, the protein(s) in the gel were transferred to a PVDF membrane using a Mini-
492 PROTEAN Tetra Cell (Bio-Rad, United States). After blocking the membrane with 3% BSA at
493 4°C overnight, the membrane was incubated with mouse anti-His6 antiserum (Santacruz Biotech,
494 USA) for 4 hours at a dilution of 1:2000. The membrane was washed once with PBS and twice
495 with PBS-Tween, followed by incubation with IgG HRP-conjugated antibody at a dilution of
496 1:10000 for 45 minutes at 4 °C. The blot was developed by using 3, 3'-Diaminobenzidine (DAB-
497 H_2O_2) as the substrate.

498

499

500 **4.3 Assay for the esterase activity of LysinB**

501 The micro-titre plate-based p-Nitrophenol butyrate (pNPB) esterase assay (Fraga *et al.*, 2019) was
502 performed to estimate the activity of LysinB_MF and LysinB_MF**Δ**PGBD. The reaction
503 conditions for both the recombinant proteins were kept constant: 5 µg of enzyme 10 mM pNPB in
504 25 mM Tris buffer (pH 7.2) in a reaction volume of 200 µl at 37ºC for 30 minutes. Absorbance at
505 410 nm was measured to determine the release of p-nitrophenol (PNP). In the control reaction,
506 enzyme buffer was added in place of the enzymes. The experiment was done twice in three sets,
507 and the absorbance value (A_{410}) was plotted as mean ± SD.

508

509 **4.4 Lipolytic activity of LysinB**

510 The enzyme activity was also observed by $CaCl_2$ precipitation. Five µg each of LysinB_MF and
511 LysinB_MF**Δ**PGBD was spot-tested on Middlebrook 7H10 plates containing 1mM $CaCl_2$ and 1%
512 (v/v) Tween-20. Precipitation, confirming lipolytic activity, was observed with 10 min incubation
513 at room temperature.

514

515 **4.5 Spot Test of LysinB**

516 For the antimycobacterial activity, the spot test on bacterial lawns was performed. Briefly,
517 *Mycobacterium smegmatis* strain mc^2155 and *M. fortuitum* (NIHJ 1615) were grown for 48 hours
518 at 37°C using Middlebrook 7H9 broth (HiMedia, India), supplemented with 0.4% glycerol.
519 Bacterial lawn in 7H10 Middlebrook agar plates containing 0.05% Tween 80 plates were prepared
520 by spreading 0.5 mL of *M. smegmatis* and *M. fortuitum* using 7H10 soft agar. Five µg each of
521 LysinB_MF and LysinB_MF**Δ**PGBD were added to the prepared lawn as spots and the plates were
522 incubated at 37°C.

Acknowledgement

We thank the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), India, for supporting the project (SERB PROJECT EMR/2017/004051). Acharya Narendra Dev College (ANDC), University of Delhi, is acknowledged for providing the infrastructural facilities and ELITE fellowship to RD. KN received the Innovation in Science Pursuit for Inspired Research (INSPIRE) fellowship from the Department of Science and Technology (DST), India; RA received the Senior Research Fellowship (SRF) from the University Grants Commission (UGC), New Delhi, India. The graphical abstract was created using BioRender.com (https://biorender.com).

ORCID

RD: 0000-0003-3171-2011

RA: 0000-0001-7666-7790

KN: 0009-0007-8807-730X

UB: 0000-0003-2441-6089

Author Contribution

RD: Data curation, Software, Visualization, Investigation, Validation, Writing- original draft. KN: Data curation, Experimental Investigation, Visualization, Validation. RA: Primer-designing, Data curation, Visualization. UB: Conceptualization, Methodology, Experimental Investigation Supervision, Validation, Writing– review & editing.

Funding

This study was supported by the funding provided to the project investigator UB by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), India

(SERB PROJECT EMR/2017/004051).

Availability of data and materials

This published article and its supplementary information files include all data generated or analyzed during this study.

Ethics approval and consent to participate.

Not applicable.

Competing interests

The authors declare that they have no competing interests.

557 **Supplementary Information**

558

559 **Table S1:** Genome annotation of phiE1336.

560

Gene	Start	Stop	Strand	Annotation	Domain
gp1	1	2469	+	Phage tail protein	-
gp2	2595	3632	+	Phage tail protein	-
gp3	3654	3821	+	HP	-
gp4	3825	4187	+	HP	-
gp5	4376	5575	+	LysinA	Amidase/PGRP
gp6	6065	5772	-	HP	-
gp7	6081	6191	+	HP	-
gp8	6188	7147	+	LysinB_MF	PG binding and AB_hydrolase
gp9	7147	7236	+	HP	-
gp10	7242	7457	+	Holin	-
gp11	7454	7900	+	HP	-
gp12	8509	8093	-	HP	-
gp13	8899	8513	-	HP	-

561

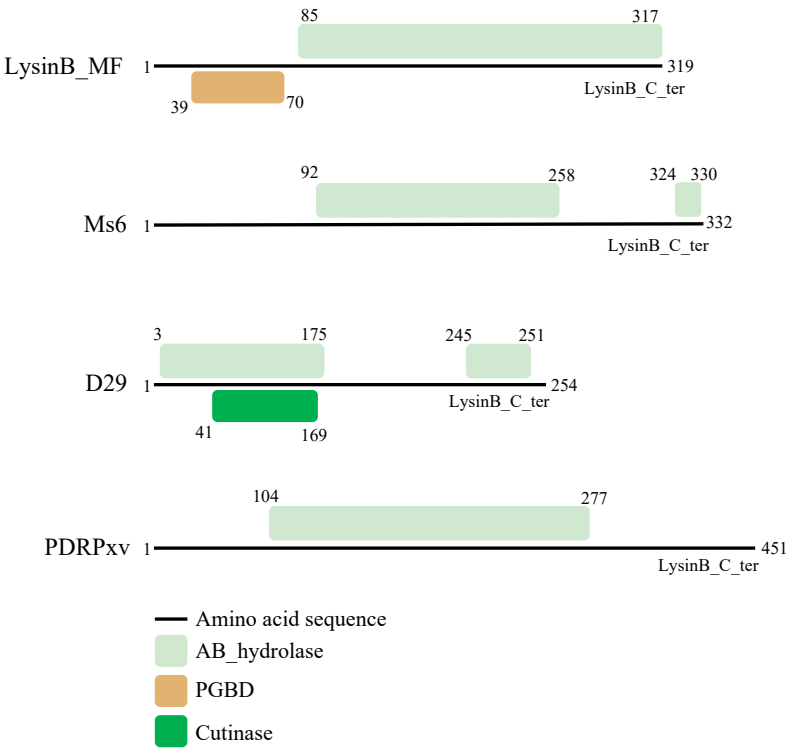


Figure S1: Comparison of domain organization of LysinBs. Domain prediction and architecture of well-studied LysinBs indicate the modular nature of LysinB_MF. Unlike other LysinBs, this lysin has a PGBD in the N-terminal end of the enzyme.

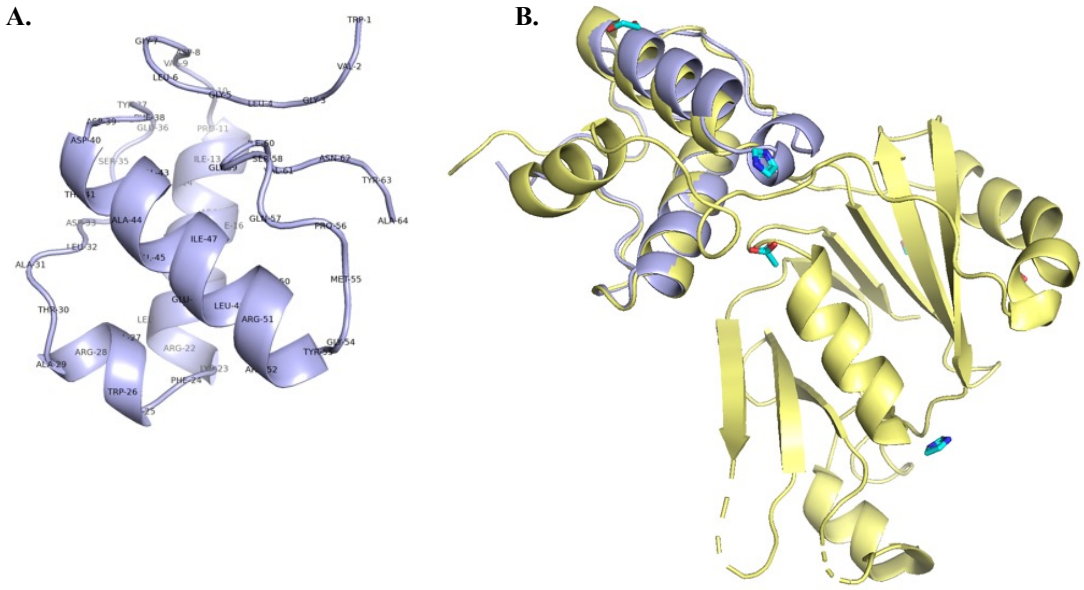


Figure S2: Structural prediction of PGBD from LysinB_MF and alignment with an amidase. A) To facilitate our understanding of the structural similarity of PGBD from LysinB_MF, we separately built its 3D model using the structure of the lysin as a template. The structure retains its original α-helices, as observed in the native LysinB_MF. However, the single β-strand is missing. The structure shows very high confidence and was taken forward for further analysis. B) A zinc-dependent amidase (PDB accession 4XXT), a peptidoglycan-binding domain-containing protein from *Clostridium acetobutylicum* (ATCC 824), displays high structural similarity with the PGBD from LysinB_MF. The two structures can be superimposed and superpositioned with an overall RMSD as low as 0.46.

579 ## References
580

581 **Abad L, Gauthier CH, Florian I, Jacobs-Sera D, Hatfull GF. 2023.** The heterogeneous and
582 diverse population of prophages in Mycobacterium genomes. *mSystems* **8**(5): e0044623.
583 **Abouhmad A, Korany AH, Grey C, Dishisha T, Hatti-Kaul R. 2020.** Exploring the Enzymatic
584 and Antibacterial Activities of Novel Mycobacteriophage Lysin B Enzymes. *Int J Mol Sci*
585 **21**(9).
586 **Aliashkevich A, Cava F. 2022.** LD-transpeptidases: the great unknown among the peptidoglycan
587 cross-linkers. *Febs Journal* **289**(16): 4718-4730.
588 **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search
589 tool. *J Mol Biol* **215**(3): 403-410.
590 **Bagcchi S. 2023.** WHO's Global Tuberculosis Report 2022. *Lancet Microbe* **4**(1): e20.
591 **Basu MK, Poliakov E, Rogozin IB. 2009.** Domain mobility in proteins: functional and
592 evolutionary implications. *Brief Bioinform* **10**(3): 205-216.
593 **Bord S, Horner A, Hembry RM, Compston JE. 1998.** Stromelysin-1 (MMP-3) and stromelysin-
594 2 (MMP-10) expression in developing human bone: potential roles in skeletal
595 development. *Bone* **23**(1): 7-12.
596 **Bradford MM. 1976.** A rapid and sensitive method for the quantitation of microgram quantities
597 of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.
598 **Cote J, Welch C, Kimble M, Archambault D, Ross JC, Orellana H, Amero K, Bourett C,**
599 **Daigle A, Hutchison KW, et al. 2022.** Characterization of the cluster MabR prophages of
600 Mycobacterium abscessus and Mycobacterium chelonae. *G3 (Bethesda)* **12**(9).
601 **Das R, Bajpai U. 2023.** Functional characterization of a DNA-dependent AAA ATPase in a F-
602 cluster mycobacteriophage. *Virus Res* **323**: 198957.
603 **Das R, Eniyan K, Bajpai U. 2021.** Computational identification and characterization of antigenic
604 properties of Rv3899c of Mycobacterium tuberculosis and its interaction with human
605 leukocyte antigen (HLA). *Immunogenetics* **73**(5): 357-368.
606 **Eniyan K, Sinha A, Ahmad S, Bajpai U. 2020.** Functional characterization of the endolysins
607 derived from mycobacteriophage PDRPxv. *World J Microbiol Biotechnol* **36**(6): 83.
608 **Fraga AG, Trigo G, Murthy RK, Akhtar S, Hebbur M, Pacheco AR, Dominguez J, Silva-**
609 **Gomes R, Goncalves CM, Oliveira H, et al. 2019.** Antimicrobial activity of
610 Mycobacteriophage D29 Lysin B during Mycobacterium ulcerans infection. *PLoS Negl*
611 *Trop Dis* **13**(8): e0007113.
612 **Fraga K, Maireles M, Jordan M, Soldevila L, Murillo O. 2022.** Mycobacterium fortuitum
613 osteomyelitis of the cuboid bone treated with CERAMENT G and V: a case report. *J Bone*
614 *Jt Infect* **7**(4): 163-167.
615 **Gigante AM, Hampton CM, Dillard RS, Gil F, Catalao MJ, Moniz-Pereira J, Wright ER,**
616 **Pimentel M. 2017.** The Ms6 Mycolyl-Arabinogalactan Esterase LysB is Essential for an
617 Efficient Mycobacteriophage-Induced Lysis. *Viruses* **9**(11).
618 **Gigante AM, Olivenca F, Catalao MJ, Leandro P, Moniz-Pereira J, Filipe SR, Pimentel M.**
619 **2021.** The Mycobacteriophage Ms6 LysB N-Terminus Displays Peptidoglycan Binding
620 Affinity. *Viruses* **13**(7).
621 **Gordon SV, Parish T. 2018.** Microbe Profile: Mycobacterium tuberculosis: Humanity's deadly
622 microbial foe. *Microbiology (Reading)* **164**(4): 437-439.
623 **Grant JR, Stothard P. 2008.** The CGView Server: a comparative genomics tool for circular
624 genomes. *Nucleic Acids Res* **36**(Web Server issue): W181-184.

625 **Hatfull GF. 2018.** Mycobacteriophages. *Microbiol Spectr* **6**(5).
626 **Hatfull GF. 2022.** Mycobacteriophages: From Petri dish to patient. *PLoS Pathog* **18**(7): e1010602.
627 **Ho YS, Adroub SA, Aleisa F, Mahmood H, Othoum G, Rashid F, Zaher M, Ali S, Bitter W,**
628 **Pain A, et al. 2012.** Complete genome sequence of Mycobacterium fortuitum subsp.
629 fortuitum type strain DSM46621. *J Bacteriol* **194**(22): 6337-6338.
630 **Holm L. 2022.** Dali server: structural unification of protein families. *Nucleic Acids Res* **50**(W1):
631 W210-W215.
632 **Hurst-Hess K, Walz A, Yang Y, McGuirk H, Gonzalez-Juarrero M, Hatfull GF, Ghosh P,**
633 **Ojha AK. 2023.** Intrapulmonary Treatment with Mycobacteriophage LysB Rapidly
634 Reduces Mycobacterium abscessus Burden. *Antimicrob Agents Chemother* **67**(6):
635 e0016223.
636 **Jacobs-Sera D, Marinelli LJ, Bowman C, Broussard GW, Guerrero Bustamante C, Boyle**
637 **MM, Petrova ZO, Dedrick RM, Pope WH, Science Education Alliance Phage Hunters**
638 **Advancing G, et al. 2012.** On the nature of mycobacteriophage diversity and host
639 preference. *Virology* **434**(2): 187-201.
640 **Khosravi AD, Mirsaeidi M, Farahani A, Tabandeh MR, Mohajeri P, Shoja S, Hoseini Lar**
641 **KhosroShahi SR. 2018.** Prevalence of nontuberculous mycobacteria and high efficacy of
642 d-cycloserine and its synergistic effect with clarithromycin against Mycobacterium
643 fortuitum and Mycobacterium abscessus. *Infect Drug Resist* **11**: 2521-2532.
644 **Knauper V, Lopez-Otin C, Smith B, Knight G, Murphy G. 1996.** Biochemical characterization
645 of human collagenase-3. *J Biol Chem* **271**(3): 1544-1550.
646 **Korany AH, Abouhmad A, Bakeer W, Essam T, Amin MA, Hatti-Kaul R, Dishisha T. 2019.**
647 Comparative Structural Analysis of Different Mycobacteriophage-Derived
648 Mycolylarabinogalactan Esterases (Lysin B). *Biomolecules* **10**(1).
649 **Korndorfer IP, Danzer J, Schmelcher M, Zimmer M, Skerra A, Loessner MJ. 2006.** The
650 crystal structure of the bacteriophage PSA endolysin reveals a unique fold responsible for
651 specific recognition of Listeria cell walls. *J Mol Biol* **364**(4): 678-689.
652 **Laemmli UK. 1970.** Cleavage of structural proteins during the assembly of the head of
653 bacteriophage T4. *Nature* **227**(5259): 680-685.
654 **Letunic I, Bork P. 2021.** Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
655 display and annotation. *Nucleic Acids Res* **49**(W1): W293-W296.
656 **Low LY, Yang C, Perego M, Osterman A, Liddington RC. 2005.** Structure and lytic activity of
657 a Bacillus anthracis prophage endolysin. *J Biol Chem* **280**(42): 35433-35439.
658 **Mishra LS, Kim SY, Caddell DF, Coleman-Derr D, Funk C. 2021.** Loss of Arabidopsis matrix
659 metalloproteinase-5 affects root development and root bacterial communities during
660 drought stress. *Physiol Plant* **172**(2): 1045-1058.
661 **Nielsen JE, Beier L, Otzen D, Borchert TV, Frantzen HB, Andersen KV, Svendsen A. 1999.**
662 Electrostatics in the active site of an alpha-amylase. *Eur J Biochem* **264**(3): 816-824.
663 **Payne K, Sun Q, Sacchettini J, Hatfull GF. 2010.** Mycobacteriophage Lysin B is a novel
664 mycolylarabinogalactan esterase. *Molecular Microbiology*.
665 **Payne KM, Hatfull GF. 2012.** Mycobacteriophage endolysins: diverse and modular enzymes
666 with multiple catalytic activities. *PLoS One* **7**(3): e34052.
667 **Pollenz RS, Bland J, Pope WH. 2022.** Bioinformatic characterization of endolysins and holin-
668 like membrane proteins in the lysis cassette of phages that infect Gordonia rubripertincta.
669 *PLoS One* **17**(11): e0276603.

670 **Porollo A, Meller J. 2010.** POLYVIEW-MM: web-based platform for animation and analysis of
671 molecular simulations. *Nucleic Acids Res* **38**(Web Server issue): W662-666.
672 **Qiu S, Azofra LM, MacFarlane DR, Sun C. 2018.** Hydrogen bonding effect between active site
673 and protein environment on catalysis performance in H2-producing [NiFe] hydrogenases.
674 *Physical Chemistry Chemical Physics*.
675 **Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.**
676 InterProScan: protein domains identifier. *Nucleic Acids Res* **33**(Web Server issue): W116-
677 120.
678 **Robert X, Gouet P. 2014.** Deciphering key features in protein structures with the new ENDscript
679 server. *Nucleic Acids Res* **42**(Web Server issue): W320-324.
680 **Schmitz JE, Ossiprandi MC, Rumah KR, Fischetti VA. 2011.** Lytic enzyme discovery through
681 multigenomic sequence analysis in Clostridium perfringens. *Appl Microbiol Biotechnol*
682 **89**(6): 1783-1795.
683 **Sethi S, Arora S, Gupta V, Kumar S. 2014.** Cutaneous Mycobacterium fortuitum Infection:
684 Successfully Treated with Amikacin and Ofloxacin Combination. *Indian J Dermatol* **59**(4):
685 383-384.
686 **Shah S, Das R, Chavan B, Bajpai U, Hanif S, Ahmed S. 2023.** Beyond antibiotics: phage-
687 encoded lysins against Gram-negative pathogens. *Front Microbiol* **14**: 1170418.
688 **Singh AK, Gangakhedkar R, Thakur HS, Raman SK, Patil SA, Jain V. 2023.**
689 Mycobacteriophage D29 Lysin B exhibits promising anti-mycobacterial activity against
690 drug-resistant Mycobacterium tuberculosis. *Microbiol Spectr* **11**(6): e0459722.
691 **Suzuki S, Yoshikawa M, Imamura D, Abe K, Eichenberger P, Sato T. 2020.** Compatibility of
692 Site-Specific Recombination Units between Mobile Genetic Elements. *iScience* **23**(1):
693 100805.
694 **Thompson JD, Higgins DG, Gibson TJ. 1994.** CLUSTAL W: improving the sensitivity of
695 progressive multiple sequence alignment through sequence weighting, position-specific
696 gap penalties and weight matrix choice. *Nucleic Acids Res* **22**(22): 4673-4680.
697 **Varadi M, Bertoni D, Magana P, Paramval U, Pidruchna I, Radhakrishnan M, Tsenkov M,**
698 **Nair S, Mirdita M, Yeo J, et al. 2023.** AlphaFold Protein Structure Database in 2024:
699 providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*.
700 **Zheng ZL, Ye MQ, Zuo ZY, Liu ZG, Tai KC, Zou GL. 2006.** Probing the importance of
701 hydrogen bonds in the active site of the subtilisin nattokinase by site-directed mutagenesis
702 and molecular dynamics simulation. *Biochem J* **395**(3): 509-515.
703 **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011.** PHAST: a fast phage search tool.
704 *Nucleic Acids Res* **39**(Web Server issue): W347-352.
705